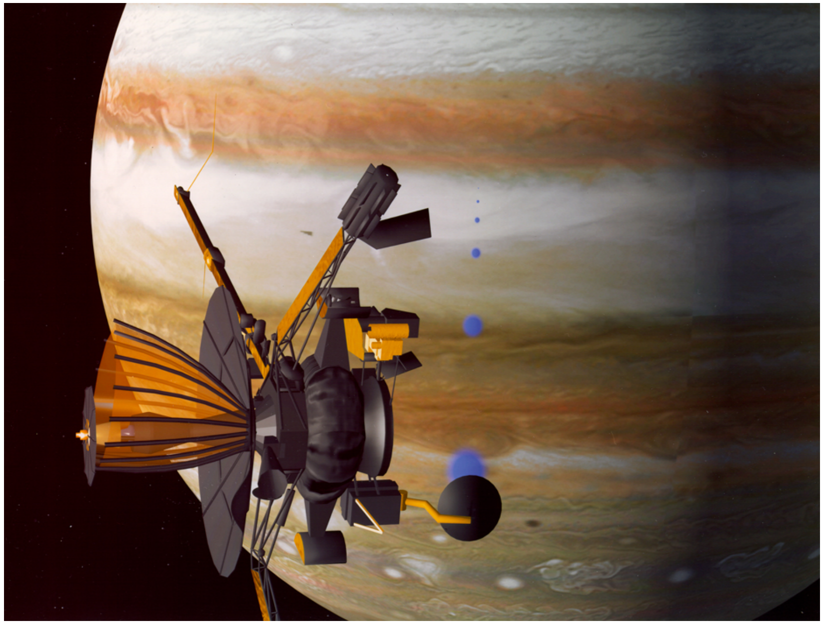


# Probabilistic Instrument Data Generation Analysis Using a Variant of Saddle-Point Approximation

Kar-Ming Cheung

5th International Planetary Probe Workshop

June 23-29, 2007, Bordeaux, France

# Motivation

- Challenges of probe missions
  - Probe spacecraft are constrained in mass and power
  - Missions are usually short and dynamic, with much uncertainties in instrument data generation and communication data return
- When considering end-to-end data delivery from instruments to Earth, other than emphasizing building better instruments and miniaturizing the instruments, one also need to consider reducing the mass and power of probe subsystems including onboard memory, data bus, and communication subsystems (antenna, gimbal, transponder, etc.)
- This paper describes a probabilistic data generation analysis method to support
  - Probe communication system design
  - Spacecraft bus, buffer, and storage design

# Problem statement (1)

- Probes are usually equipped with multiple sensor instruments with different data generation rates
- To reduce bandwidth, data compression is usually employed
  - Data compression is a fixed-to-variable data conversion process depending on the data content and the algorithm used, thus data are generated in a non-deterministic way
  - Prior probe mission experience indicates that data generation statistics are usually non-Gaussian
- Shortcomings of traditional analysis approaches
  - Simulation:
    - Time consuming
    - Lack of good random number generators for some probability distribution functions (pdf) can contribute to misleading and erroneous conclusions
  - Analysis based on Gaussian assumption
    - Gaussian assumption is not valid for a small number of instruments (< 30)

# Problem statement (2)

- In this paper, we apply a variant of Saddle-Point approximation technique [1] to estimate the tail probability of the aggregate statistics of instrument data
  - Does not resort to the Gaussian assumption
  - Improve the accuracy of estimating instrument data generation
  - In turn provides accurate estimates to optimize spacecraft design in the areas of onboard storage, bus bandwidth, downlink capabilities, etc.
- This technique can also be applied to a wide range of engineering problems including link analysis, queue overflow probabilities for different traffic models, and mass, power, cost estimations for spacecraft design, and reliability analysis

[1] C. W. Helstrom, "Approximate Evaluation of Detection Probabilities in Radar and Optical Communications," IEEE Trans. Aerosp. Electron. Syst.; vol 14, pp.630-640, July 1978.

# Problem statement (3)

- Some notations
  - $x_1, x_2, \ldots x_n$ are $n$ independent random variable with pdf $f_{xi}(x_i)$
  - $z$ is the sum of $x_1, x_2, \ldots x_n$
  - $\Psi_{xi}(s)$ is the characteristic function of $x_i$, and $\Psi_z(s)$ is the characteristic function of $z$
  - $q_+(\alpha)$ is the tail probability of $z$

$$z = \sum_{i=1}^{n} x_i \qquad f_z(z) = f_{x_1}(x_1) * f_{x_2}(x_2) * \ldots * f_{x_n}(x_n)$$

$$\Psi_z(s) = \int_{-\infty}^{\infty} e^{sz} f_z(z) dz \qquad \Psi_z(s) = \Psi_{x_1}(s) \Psi_{x_2}(s) \cdots \Psi_{x_n}(s)$$

$$q_+(\alpha) = \int_{\alpha}^{\infty} f_z(z) dz \qquad \boxed{q_+(\alpha) \approx \frac{e^{\psi(s_0)}}{\sqrt{2\pi\psi''(s_0)}}}$$

- Analysis challenges
  - Evaluation of pdf of sum of $n$ variables requires $n$-$1$ nested integration
  - Inverse of $\Psi_z(s)$ is usually extremely difficult, if not impossible

# Tables of $\Psi_z(s)$ for some popular pdf

| NAME | PROBABILITY DENSITY FUNCTION (PDF) | CHARACTERISTIC FUNCTION $\Psi(S)$ |
|---|---|---|
| Uniform distribution (continuous) | $\frac{1}{b-a}$ for $\alpha \le x \le b$<br><br>$0$ for $x < a$ or $x > b$ | $\frac{e^{sb} - e^{sa}}{s(b-a)}$ |
| Triangular distribution | $\frac{2(x-a)}{(b-a)(c-a)}$ for $\alpha \le x \le c$<br>$\frac{2(b-x)}{(b-a)(b-c)}$ for $c \le x \le b$ | $2\frac{(b-c)e^{as} - (b-a)e^{as} + (c-a)e^{bs}}{(b-\alpha)(c-a)(b-c)s^2}$ |
| Normal distribution | $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ | $e^{\left(\mu s + \frac{\sigma^2 s^2}{2}\right)}$ |
| Exponential distribution | $\lambda e^{-\lambda x}$ | $\left(1 - \frac{s}{\lambda}\right)^{-1}$ |
| Gamma distribution | $x^{k-1}\frac{e^{(-x/\theta)}}{\Gamma(k)\theta^k}$ | $(1-\theta s)^{-k}$ for $s < 1/\theta$ |
| Beta distribution | $\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ | $1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{s^k}{k!}$ |

# Outline of Approx Term Derivation (1)

1. $\Psi_z(s)$ as Laplace Transform of $f_z(z)$ $\quad F_z(s) = L\{f_z(u)\} = \Psi_z(-s)$

2. $f_z(z)$ as inverse Laplace Transform of $\Psi_z(s)$ $\quad f_z(z) = \frac{1}{2\pi j}\int_{\sigma'-j\infty}^{\sigma'+j\infty} \Psi_z(-s)e^{sz}ds$

3. Express $q_+(\alpha)$ using 2. $\quad q_+(\alpha) = \int_\alpha^\infty f_z(z)dz = \int_\alpha^\infty \frac{1}{2\pi j}\int_{\sigma'-j\infty}^{\sigma'+j\infty} \Psi_z(-s)e^{sz}dsdz$

$$= \frac{1}{2\pi j}\int_{\sigma'-j\infty}^{\sigma'+j\infty} \Psi_z(-s)\int_\alpha^\infty e^{sz}dzds$$

$$= \frac{1}{2\pi j}\int_{\sigma'-j\infty}^{\sigma'+j\infty} \Psi_z(-s)\frac{\left[e^{sz}\right]_\alpha^\infty}{s}ds$$

# Outline of Approx Term Derivation (2)

4. For $\sigma' < 0$, we have $$q_+(\alpha) = \frac{-1}{2\pi j}\int_{\sigma'-j\infty}^{\sigma'+j\infty} \Psi_z(-s)\frac{e^{s\alpha}}{s}\,ds$$

5. Change of variable $s \leftrightarrow -s$, and define $\sigma = -\sigma'$ $$q_+(\alpha) = \frac{-1}{2\pi j}\int_{-\sigma'+j\infty}^{-\sigma'-j\infty} \Psi_z(s)\frac{e^{-s\alpha}}{s}\,d(-s)$$ $$= \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} \Psi_z(s)\frac{e^{-s\alpha}}{s}\,ds$$

6. Define $\psi(s)$ as $$e^{\psi(s)} = \frac{e^{-s\alpha}\Psi_z(s)}{s}$$

7. Denote $s_o$ such that $\psi'(s_0) = 0$

8. $\psi(s)$ can be approximate as (using Taylor Series expansion) $$\psi(s) \approx \psi(s_0) + \frac{1}{2}\psi''(s_0)(s-s_0)^2$$

# Outline of Approx Term Derivation (3)

9. Substitute (6) (8) in (5) and let $\sigma = s_o$

$$q_+(\alpha) = \frac{1}{2\pi j}\int_{s_0 - j\infty}^{s_0 + j\infty} e^{\psi(s)} ds$$
$$\approx \frac{1}{2\pi j}\int_{s_0 - j\infty}^{s_0 + j\infty} e^{\psi(s_0) + \frac{1}{2}\psi''(s_0)(s - s_0)^2} ds$$
$$= \frac{e^{\psi(s_0)}}{2\pi j}\int_{s_0 - j\infty}^{s_0 + j\infty} e^{\frac{1}{2}\psi''(s_0)(s - s_0)^2} ds$$

10. Change of variable $s \leftrightarrow s_0 + jy$

$$q_+(\alpha) = \frac{e^{\psi(s_0)}}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\psi''(s_0) y^2} dy = \frac{e^{\psi(s_0)}}{\sqrt{2\pi\psi''(s_0)}}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi / \psi''(s_0)}} e^{-\frac{1}{2}\psi''(s_0) y^2} dy$$

11. RHS = 1, and

$$\boxed{q_+(\alpha) \approx \frac{e^{\psi(s_0)}}{\sqrt{2\pi\psi''(s_0)}}}$$

- A more accurate approximation

$$q_+(\alpha) \approx \frac{e^{\psi(s_0)}}{\sqrt{2\pi\psi''(s_0)}}\left(1 + \frac{\psi^{(4)}(s_0)}{8[\psi''(s_0)]^2}\right)$$

# Computation Procedure

1. Evaluate $\Psi_z(s)$ $$\Psi_z(s)=\Psi_{x_1}(s)\,\Psi_{x_2}(s)\cdots\Psi_{x_n}(s)$$

2. Construct $\psi(s)$ $$\psi(s)=-s\alpha+Log\,\Psi_z(s)-Log\,s$$

3. Construct $\psi'(s)$ $$\psi'(s)=-\alpha+\frac{\Psi'_z(s)}{\Psi_z(s)}-\frac{1}{s}$$

4. Construct $\psi''(s)$ $$\psi''(s)=\frac{\Psi''_z(s)}{\Psi_z(s)}-\frac{\Psi'_z(s)^2}{\Psi_z(s)^2}+\frac{1}{s^2}$$

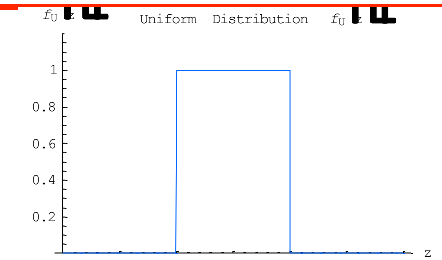
5. Find $s_o$ such that $\psi'(s_o)=0$

6. Evaluate $q_+(\alpha)$ $$\boxed{q_+(\alpha)\approx\frac{e^{\psi(s_0)}}{\sqrt{2\pi\psi''(s_0)}}}$$

# Computing $q_+(\alpha)$ of Some Simple pdf

Uniform distribution: $f_U(z) = 1$ for $1 \le z \le 2$

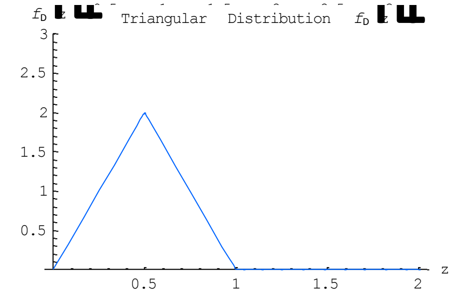
$= 0$ for $z < 1$ or $z > 2$



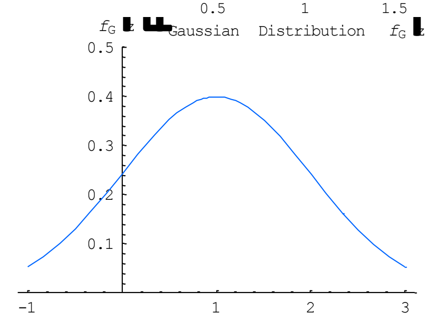

Triangular distribution: $f_\Delta(z) = 4z$ for $0 \le z \le 0.5$

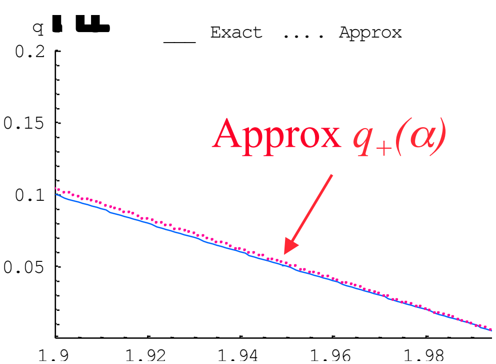
$= 4(1-z)$ for $0.5 < z \le 1.0$

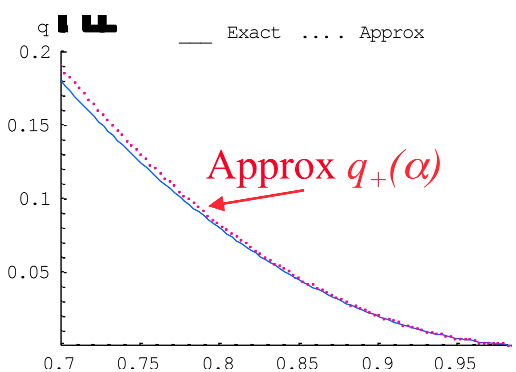
$= 0$ for $z < 0$ or $z > 1$

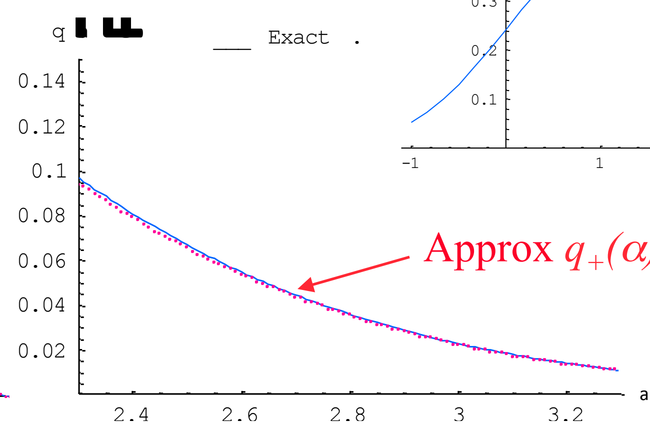



Gaussian distribution: $f_G(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{(z-1)^2}{2}}$ for $-\infty \le z \le \infty$



# A Hypothetical Probe Instrument Data Generation Examples

- Consider a probe carries 3 instruments A, B, and C; all employ data compression to reduce data generation bandwidth
- Instantaneous data rate (in unit of kilo-bits per second) of each instrument is characterized statistically by a triangular distribution with the following parameters (Table on p.6)

| | min ( $a$ ) | max ( $b$ ) | mode ( $c$ ) |
|---|---|---|---|
| Instrument A | 0 | 1 | 0.5 |
| Instrument B | 2 | 7 | 2.5 |
| Instrument C | 4 | 7 | 4.5 |

- We consider two operation scenarios:
  I. when instruments A and B are turned
  II. when instruments A, B, and C are turned on

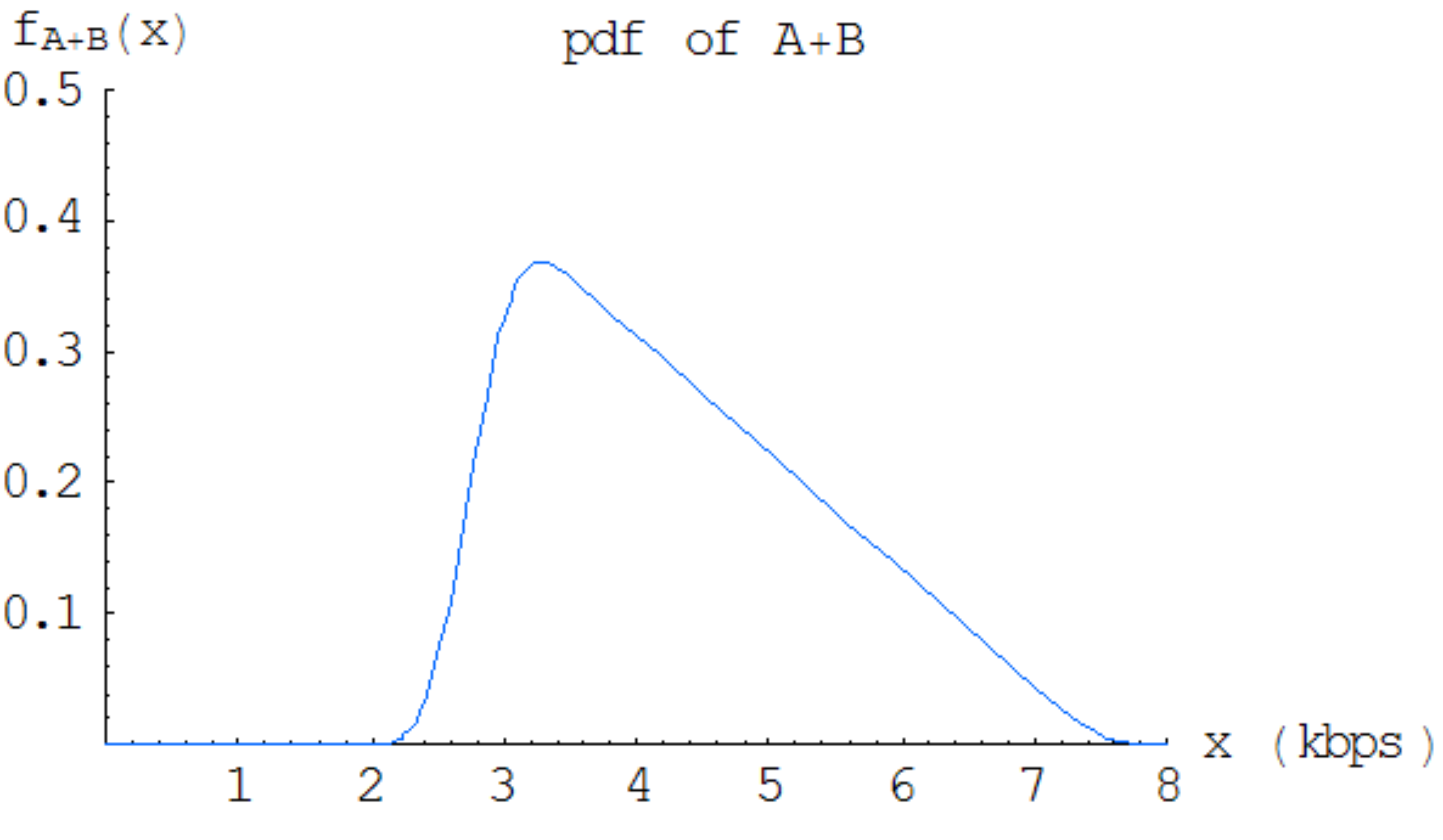
# Case I: Instrument A and B are On (1)

- The aggregate pdf is given by

$$
f_{A+B} = \begin{cases}
-\frac{8}{135}\left(8 - x\right)^4 & \frac{15}{2} < x < 8 \\
-\frac{2}{45}\left(\ldots 5 - 2x\right) & \frac{7}{2} < x \le 7 \\
-\frac{2}{135}\left(\ldots 3 - 3x\right) & x = \frac{15}{2} \\
\frac{8}{15}\left(\ldots 3 + \ldots 2x - 6x^2 + x^3\right) & 2 < x \le \frac{5}{2} \\
\frac{2}{135}\left(788 + 519x - 96x^2 + 4x^3\right) & x \text{ ä } \frac{7}{2} \\
\frac{2}{135}\left(132\ldots + 582x - 84x^2 + 4x^3\right) & 7 < x < \frac{15}{2} \\
\frac{4}{135}\left(\ldots 35\ldots 732x - 210x^2 + 20x^3\right) & 3 < x < \frac{7}{2} \\
-\frac{4}{135}\left(650 + 645x - 204x^2 + 20x^3\right) & x = 3 \\
-\frac{4}{135}\left(731 + 834x - 312x^2 + 38x^3\right) & \frac{5}{2} < x < 3
\end{cases}
$$

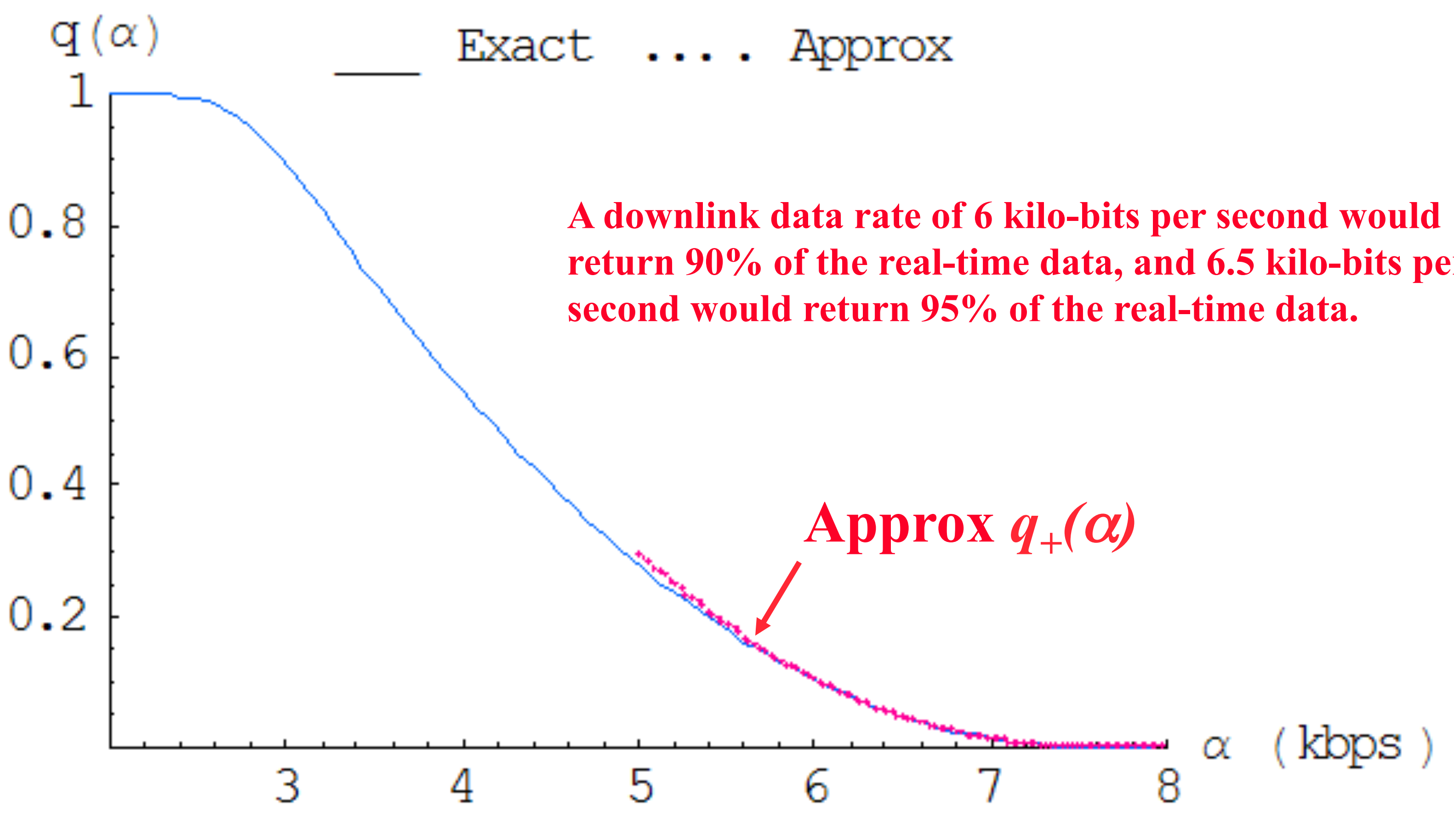
# Case I: Instrument A and B are On (2)

- Shape of $f_{A+B}(x)$



# Case I: Instrument A and B are On (3)

- Exact and approximate tail probability $q_+(\alpha)$



# Case II: Instrument A, B, and C are On (1)

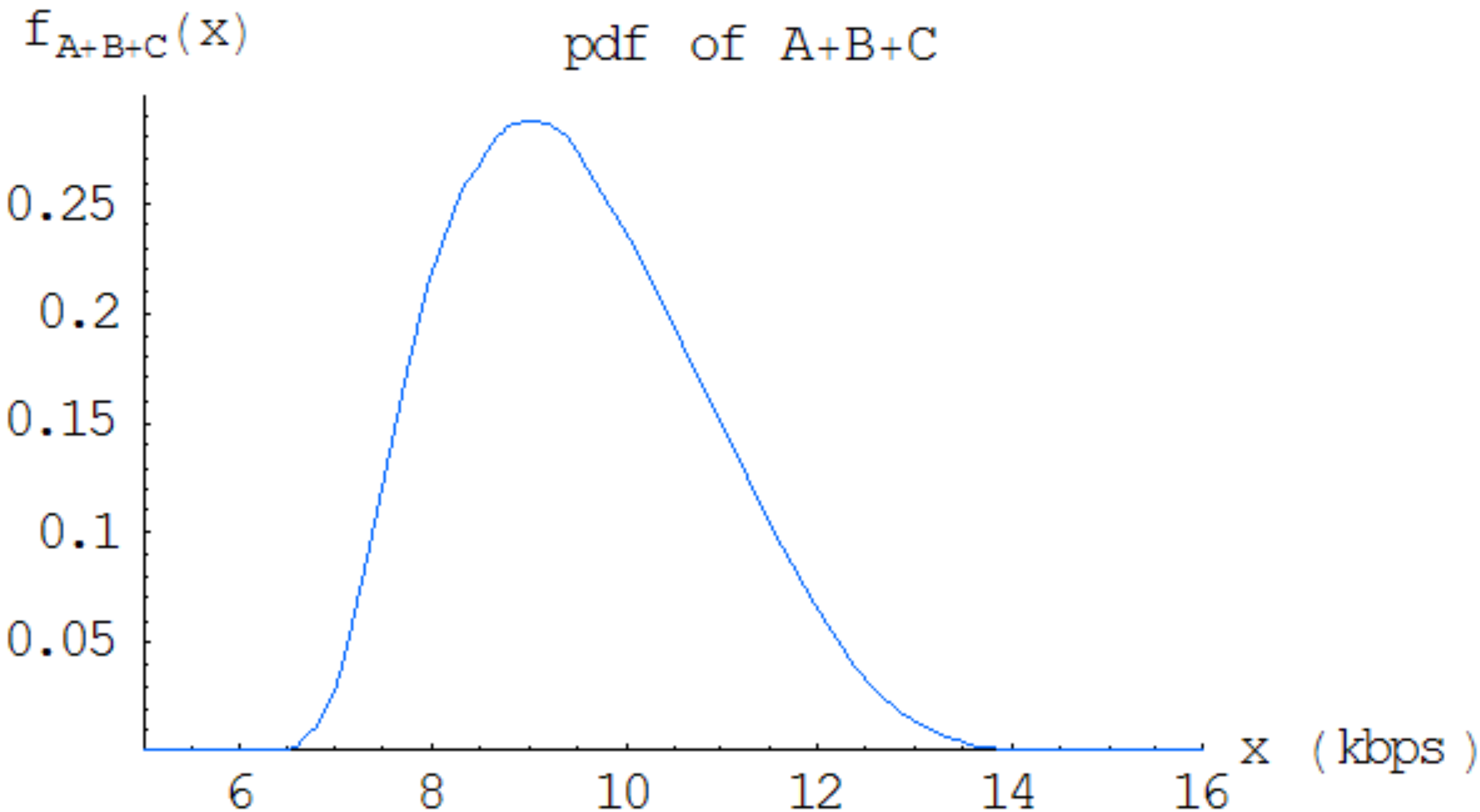
- The aggregate pdf is given by

$$
f_{A+B+C} = \begin{cases}
\frac{4503-355x}{4050} & x = \frac{23}{2} \\
\frac{2557-194x}{4050} & x = \frac{25}{2} \\
\frac{199-14x}{4050} & x = 14 \\
\frac{73-5x}{2025} & x = \frac{29}{2} \\
-\frac{8\,[\text{illegible}]}{1[\text{illegible}]} & \frac{29}{2} < x < 15 \\
\frac{8}{225}\,[\text{illegible}] & 6 < x \le \frac{13}{2} \\
-\frac{4\,[\text{illegible}]}{13[\text{illegible}]} & \frac{21}{2} < x \le 11 \\
-\frac{4\,([\text{illegible}]23[\text{illegible}]+185x)}{10125} & x = 12 \\
\frac{48807-10094x+696x^2-16x^3}{4050} & \frac{25}{2} < x < 14 \\
\frac{-32007+8654x-696x^2+16x^3}{4050} & 8 < x \le 9 \\
\frac{-330321+87620x-6960x^2+160x^3}{40500} & x = \frac{19}{2} \\
\frac{30416701-23038570x+6969160x^2-1052240x^3+79280x^4-2384x^5}{20250} & \frac{13}{2} < x < 7 \\
\frac{26002295-13477750x+2787800x^2-287600x^3+14800x^4-304x^5}{20250} & \frac{19}{2} < x < 10 \\
\frac{6259449-3923506x+982344x^2-122864x^3+7680x^4-192x^5}{4050} & \frac{15}{2} < x < 8
\end{cases}
$$

$$
f_{A+B+C} = \begin{cases}
\frac{12374493-6532250x+1338000x^2-131200x^3+6000x^4-96x^5}{20250} & x = 8 \\
\frac{4141858-2691325x+676000x^2-80600x^3+4400x^4-80x^5}{20250} & x = 7 \\
\frac{4004945-1842415x+328640x^2-28040x^3+1120x^4-16x^5}{20250} & x = 10 \\
\frac{4\,([\text{illegible}]21532+146365x-26620x^2+2420x^3-110x^4+2x^5)}{2025} & 11 < x < \frac{23}{2} \\
\frac{4\,([\text{illegible}]631605+1458535x-233940x^2+18740x^3-750x^4+12x^5)}{10125} & 12 < x < \frac{25}{2} \\
\frac{-6395073+2637585x-423360x^2+32760x^3-1200x^4+16x^5}{20250} & x = \frac{21}{2} \\
\frac{-8361149+3022810x-435560x^2+31280x^3-1120x^4+16x^5}{20250} & 14 < x < \frac{29}{2} \\
-\frac{4\,([\text{illegible}]828683+2066585x-353580x^2+30220x^3-1290x^4+22x^5)}{10125} & \frac{23}{2} < x < 12 \\
\frac{-4079541+1944450x-370440x^2+35280x^3-1680x^4+32x^5}{4050} & 10 < x < \frac{21}{2} \\
\frac{-8663091+4767190x-1053240x^2+116720x^3-6480x^4+144x^5}{20250} & 9 < x < \frac{19}{2} \\
\frac{-139899855+82846 75x-1895520x^2+206120x^3-10320x^4+176x^5}{20250} & x = \frac{15}{2} \\
\frac{-53752755+37082470x-10208280x^2+1401680x^3-96000x^4+2624x^5}{20250} & 7 < x < \frac{15}{2}
\end{cases}
$$

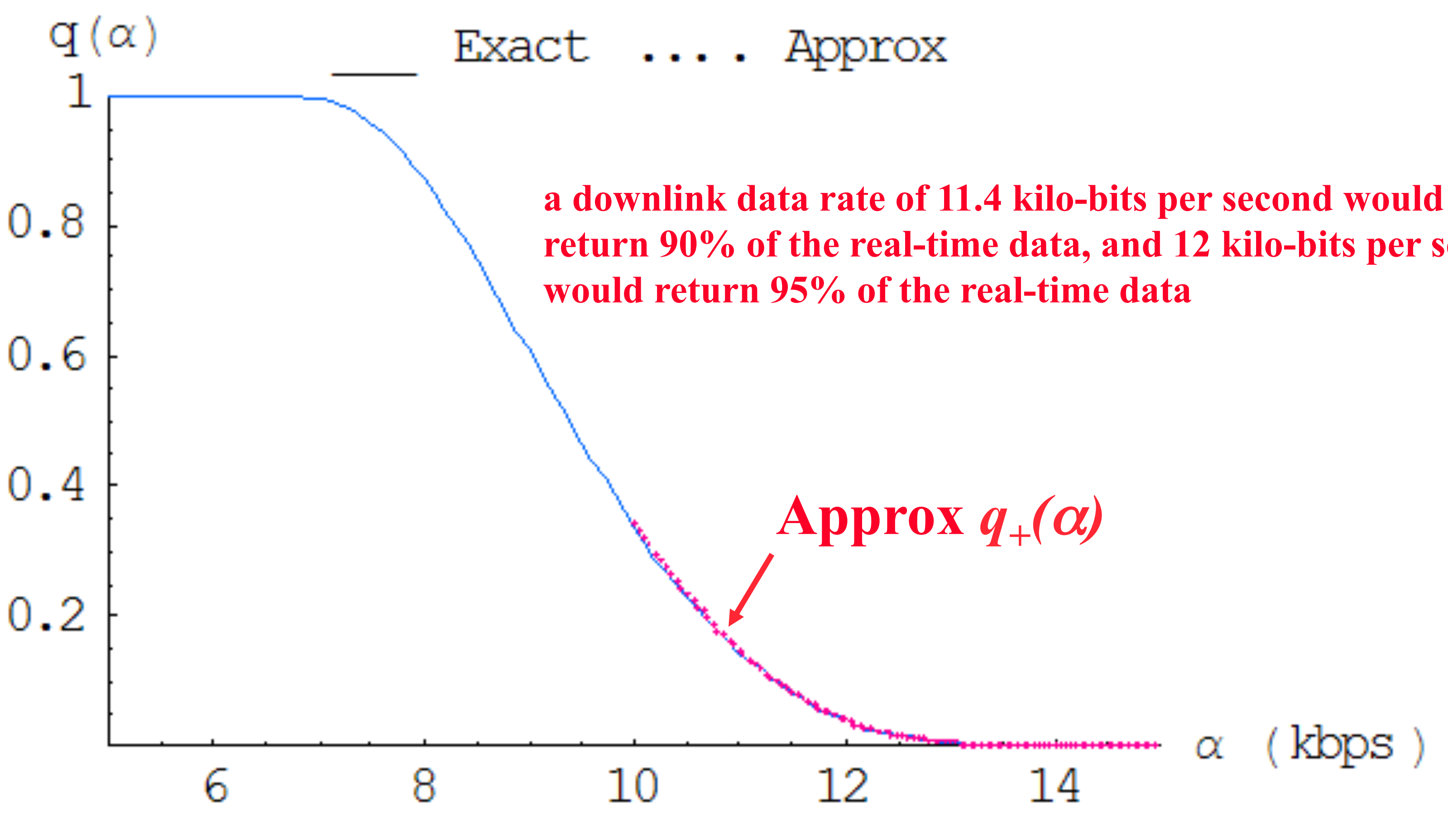
# Case II: Instrument A,B, and C are On (2)

- Shape of $f_{A+B+C}(x)$



# Case II: Instrument A, B, and C are On (3)

- Exact and approximate tail probability $q_+(\alpha)$



# Conclusion

- We demonstrate that the Saddle-Point approximation method greatly improves the accuracy of estimating instrument data generation for a mission
- This in turn provides accurate estimates to optimize spacecraft design in the areas of onboard storage, bus bandwidth, downlink capabilities, etc
- This technique can also be applied to a wide range of engineering problems, reliability analysis, and risk management problems

# Acknowledgement


- This research was funded by the Architecture Modeling and Simulation Effort of the SCaN Constellation Integration Program. The method describes in this paper has many engineering analysis applications. In this paper we apply the technique to the problem of instrument data generation of spacecraft probe

- The author would like to thank Dr. Michela Muñoz Fernández for the insightful discussion on probabilistic spacecraft data volume problem